

# Weekly Report

Lu Junhua

2016 年 7 月 4 日

## Done

- Finished a course paper, and gave a presentation on this paper & its experiment.
- Discussed about the new design of our vis paper.
- Started adding some details of vis paper; also began designing the illustrations of workflow and our model.
- *miscellaneous*: curing fever; helped girlfriend transport for her dormitory change.

## To do

- Finish revising the model part. Consulting with Prof. Peng for detailed revision advice.
- Some evaluations on new design.

## 论文

- EuroVis 15 *Small MultiPiles: Piling Time to Explore Temporal Patterns in Dynamic Networks* 文章结合了flipping or animation; small multiples和summary views in isolation这几种技术, 对密集并带权的并且是时变网络进行可视探索. 相比于之前审的call graph, 很显然这篇文章看起来更为舒心一些. 其设计初衷是帮助其他一些领域的人在进行正式的统计分析或者其他分析前能有个很不错的数据探索. 基于这个目的, 设计中仍然是以矩阵形式的图为主导的. 工作有一些简明的特点, 如: 对于时间片的划分, 是可以用户指定的, 不一定都定长, 当然也可以靠机器学习算法自动推断, 每一段都是一个pile; 设计的matrix preview是对一个pile的概览图, 感觉是一个很细致的设计; cover matrix作为一种聚合的表达方式, 支持不同的编码方法; 支持大量交互. 作为一个初步探索工具, 虽不华丽, 但我觉得已经很实用.
- Pvis 2016 *PieceStack: Toward Better Understanding of Stacked Graphs* 堆叠图是常见的时变数据表示方法, 尽管看起来美观但是往往难于比较, 并且层数不能太大, 否则视觉上非常混乱. 本文提出的方法是在原始的堆叠图基础上, 进行一些信息的补充. 补充的内容, 来自于对时变数据分析任务的深刻调研, 对于时间分隔, 时间段上的聚类聚合信息, 采用了一套处理方法, 其中提到了Douglas-Peucker算法这个新名词, 其本身比较简单应该也是时间序列中常用的一种方法. 这些聚类信息被嵌入到原始的堆叠图上, 帮助更好的理解数据, 发现一些单纯用堆叠图无法发现的模式.

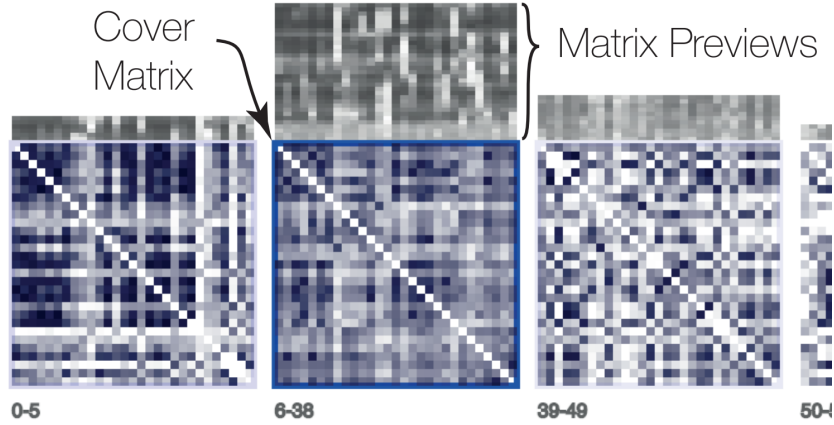


图 1: MultiPiles.

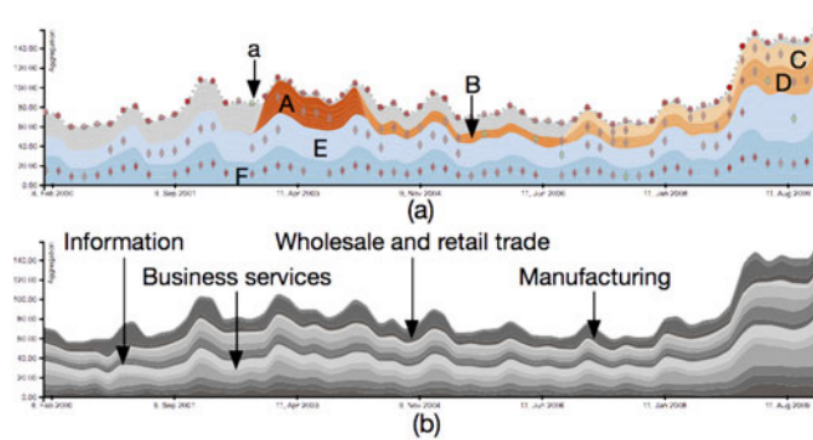


图 2: PieceStack.

- FacetAtlas: Multifaceted Visualization for Rich Text Corpora 这篇文章本来是用作文本可视化的, 但其实其设计思路是可以用到别的地方的. 比如我们就期望改进我们原来的中间视图, (算是对一个二部图的改造). 个人感觉曹老师的文章配色非常素雅, 高亮部分尽管用对比色仍然显得非常合适美观, 这一点值得学习. 另外他的布局算法也经常是比较复杂的, 比如本文中那种对力引导改进的点布局方法等等, 包括利用了一些优化算法.
- VIS16 *A Linguistic Approach to Categorical Color Assignment for Data Visualization* 文章介绍的是将语义信息和颜色编码结合起来的一种方法. 一开始我以为是一种全自动算法, 后来发现其过程仍是比较复杂的. 其中也用到了大量数据处理和挖掘. 其首先对可着色性有一个度量, 利用的是谷歌的语料库, 结合NLP那些n-gram等等的信息, 再搞了一个互信息的指标, 得到这个可着色性分数. 然后将要被编码的项名称与颜色一起拿到谷歌搜图中去进行查询. 搜索时还考虑近义词等问题, 用WordNet的语料库来加强搜索结果. 有点奇怪的是文章没有作user study之类的, 我也无法从感知上断定其优劣, 虽然在未来工作中也有提到这一点.

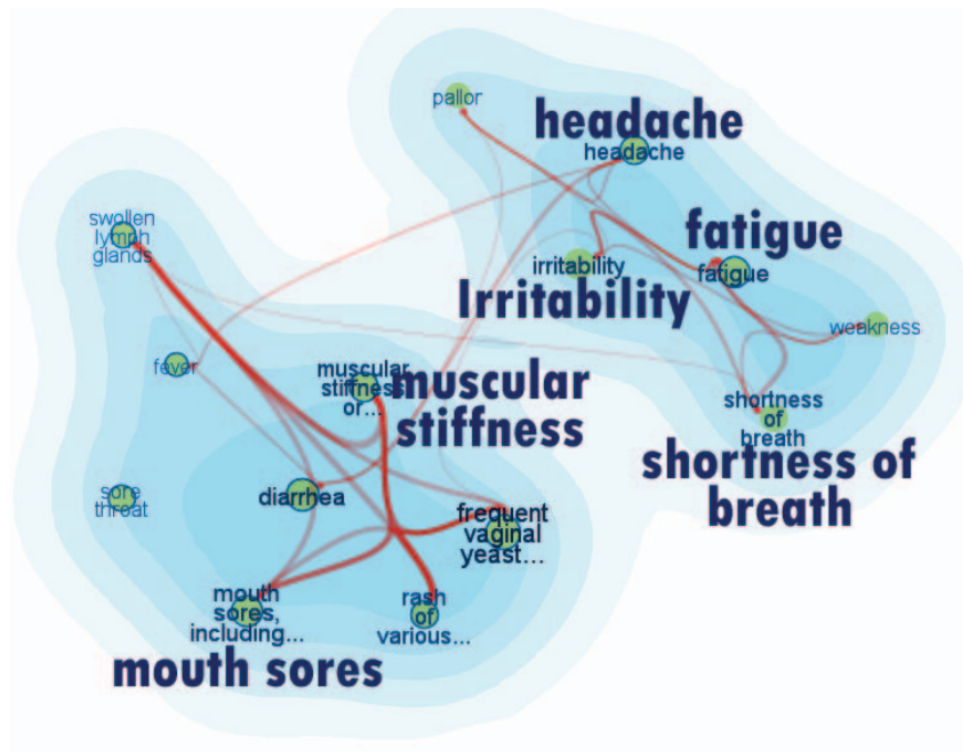


图 3: FacetAtlas.

- *Mining Graphs for Understanding Time-Varying Volumetric Data* 尽管这是关于时变体数据的图, 一些想法仍然可以拿来用, 纯粹从图的角度来说. 文中提到的三种挖掘方法, 第一种图的简化, 似乎在之前有见过, 简化为fan, connector或clique, 我想如果有别的明显的结构, 我们也可以自定义这样一些简化的方法, 但是需要有配套的检测方法. 第二种是community detection, 基于label propagation算法. 注意到community 是比clique要弱的一种一群人之间的联系, 文中使用bubble sets的可视化方法来进行展示. 第三种是可视推荐, 这个还不同于之前那篇讲可视推荐的文章(定义好坏或者说推荐度的指标比较trick), 他有一定的套路在之中, 文中又分为node recommendation和community recommendation.